AD-A033 735 MARYLAND UNIV COLLEGE PARK DEPT OF COMPUTER SCIENCE F/G 12/1
ON THE PERTURBATION OF PSEUDO-INVERSES, PROJECTIONS AND LINEAR --ETC(U)
JUL 75 G W STEWART N00014-67-A-0218-0018
UNCLASSIFIED TR-387 NL

1 OF 1
AD
A033735

END
DATE
FILMED
2-77

1.0
4.5
5.0
5.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

See 1473

12

COMPUTER SCIENCE TECHNICAL REPORT SERIES

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742

(12)

Technical Report TR-387
N00014-67-A-0218-0018

July 1975

ON THE PERTURBATION OF
PSEUDO-INVERSES, PROJECTIONS, AND
LINEAR LEAST SQUARES PROBLEMS

by

G. W. Stewart

Abstract

This paper surveys perturbation theory for the pseudo-inverse (Moore-Penrose generalized inverse), for the orthogonal projection onto the column space of a matrix, and for the linear least squares problem.

ON THE PERTURBATION OF PSEUDO-INVERSES, PROJECTIONS AND LINEAR LEAST SQUARES PROBLEMS*

G. W. Stewart**

1. Introduction

The pseudo-inverse (or Moore-Penrose generalized inverse) of a matrix A may be defined as the unique matrix $A^\dagger$ satisfying the following conditions [due to Penrose (1955)]:

(1.1a) $$A^\dagger A A^\dagger = A^\dagger ,$$

(1.1b) $$A A^\dagger A = A ,$$

(1.1c) $$(A A^\dagger)^H = A A^\dagger ,$$

(1.1d) $$(A^\dagger A)^H = A^\dagger A .$$

The pseudo-inverse and its generalizations have been extensively investigated and widely applied. One reason for this interest in the pseudo-inverse is that it permits the succinct expression of some important geometric constructions in n-dimensional space. This paper will be concerned with the pseudo-inverse and two related geometric constructions: the orthogonal projection onto a subspace and the linear least squares problem.

The orthogonal projection onto a subspace $\mathcal{X}$ is the unique Hermitian, idempotent matrix P whose column space [denoted by $\mathcal{R}(P)$] is $\mathcal{X}$. It follows

*This work was supported in part by the Office of Naval Research under Contract No. N0014-67-A-0128-0018.

**Computer Science Department, University of Maryland, College Park, Md. 20742.

from (1c) that the matrix

$$P_A = AA^{\dagger}$$

is Hermitian and from (1b) that P_A is idempotent and $R(P_A) = R(A)$. Hence P_A is the orthogonal projection onto $R(A)$. A similar argument shows that

$$R_A = A^{\dagger}A \tag{1.2}$$

is the projection onto $R(A^H)$, the row space of A.

The second construction is the solution of the linear least squares problem of choosing a vector x to minimize

$$\rho(x) = \|b-Ax\|_2 , \tag{1.3}$$

where b is a fixed vector and $\|\cdot\|_2$ denotes the usual Euclidean norm. The solutions of this problem are given by

$$x = A^{\dagger}b + (I-R_A)z , \tag{1.4}$$

where z is arbitrary. When A has full column rank, $R_A = I$ and the solution $x = A^{\dagger}b$ is unique. Otherwise, it is easily verified from (1.1) and (1.2) that $A^{\dagger}b$ is orthogonal to $(I-R_A)z$, so that by the Pythagorean theorem

$$\|x\|_2^2 = \|A^{\dagger}b\|_2^2 + \|(I-R_A)z\|_2^2 .$$

It follows that $x = A^{\dagger}b$ is the unique solution of (1.3) that has minimal norm.

The object of this paper is to describe the effects of perturbations in A on $A^{\dagger}$, on P_A and on $A^{\dagger}b$; i.e., on the pseudo-inverse, on the projection onto $R(A)$, and on the solution of the linear least squares problem. Such descriptions are important for three reasons. First the results are useful mathematical tools. Second, in numerical applications the elements of A will seldom be known exactly, and it is necessary to have bounds on the effects of the uncertainties in A. Finally many numerical processes for computing projections and least squares solutions behave as if exact computations had been performed on a perturbed matrix $A + E$, where E is a small matrix whose size depends on the algorithm and the arithmetic used in its execution.

We shall be concerned with three kinds of results: perturbation bounds, asymptotic expressions, and derivatives. The perturbation bounds are needed in the applications mentioned above. Asymptotic expressions and derivatives are useful computational tools when the perturbation is actually known. Moreover they can be used to check the sharpness of the perturbation bounds. Not surprisingly it is rather difficult to obtain a reasonably sharp perturbation bound that tells the complete story of the effects of the perturbations. Asymptotic forms and derivatives are easier to come by.

In order to make this survey reasonably self-contained, we begin in §2 with a review of the necessary background. In §3 we develop the perturbation theory for the pseudo-inverse, in §4 for the projection P_A, and in §5 for the least squares solution $A^{\dagger}b$.

Notes and references. For background on the generalized inverse see the books by Ben-Israel and Greville (1974), Boullion and Odell (1971), and Rao and Mitra (1971). The expression (1.1) is due to Penrose (1955, 1956) whose papers initiated the current interest in the pseudo-inverse.

Many articles on perturbation theory for pseudo-inverses and related problems have appeared in the literature. To date the most complete survey of the problem has been given by Wedin (1973). In addition to collecting and unifying earlier material, this paper will present some new results.

2. Preliminaries

Notation. Throughout this paper we shall use the notational conventions of Householder (1964). Specifically, matrices are denoted by upper case Latin and Greek letters, vectors by lower case Latin letters, and scalars by lower case Greek letters. The symbol $\mathbb{C}$ denotes the set of complex numbers, $\mathbb{C}^n$ the set of complex n-vectors, and $\mathbb{C}^{m\times n}$ the set of complex $m \times n$ matrices. The matrix A^H is the conjugate transpose of A. The column space of A is denoted by $\mathcal{R}(A)$, and its orthogonal complement by $\mathcal{R}(A)^\perp$.

We shall be concerned with a fixed matrix $A \in \mathbb{C}^{m\times n}$ with

$$\text{rank}\,(A) = r\,.$$

The matrix $E \in \mathbb{C}^{m\times n}$ will denote a perturbation of A and we shall set

$$B = A + E\,.$$

Since we are concerned with the geometry of $\mathbb{C}^n$, we shall be at some pains to cast our results in such a way that they are not affected by unitary transformations (cf. the section on unitarily invariant norms below). We may use this fact to transform our perturbation problems into a simpler form. Specifically, let $U = (U_1, U_2) \in \mathbb{C}^{m\times m}$ be a unitary matrix with $R(U_1) = R(A)$ and let $V = (V_1, V_2)$ be a unitary matrix with $R(V_1) = R(A^H)$. Then U^HAV has the form

$$U^HAV = \begin{pmatrix} A_{11} & 0 \\ 0 & 0 \end{pmatrix} \tag{2.1}$$

where $A_{11} \in \mathbb{C}^{r\times r}$ is nonsingular. We shall partition U^HEV and U^HBV conformally with U^HAV:

$$U^HEV = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix},$$

$$U^HBV = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}+E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}.$$

These forms will be called reduced forms of the matrices A, B, and E, and in the sequel we shall often assume that the matrices are in reduced form. In this case, the pseudo-inverse is given by

$$A^\dagger = \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.2}$$

Singular values. It is a well-known result that in the reduced form (2.1) the matrices U_1 and V_1 may be chosen so that

$$A_{11} \equiv \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r) ,$$

where

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 .$$

This reduced form is called the singular value decomposition of the matrix A, and the numbers σ_i are called the singular values of A. From the relation (2.2) and the fact that $(U^H AV)^\dagger = V^H A^\dagger U$, it follows that

$$A^\dagger = V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^H .$$

The i-th singular value of a matrix A, which will be denoted by $\sigma_i(A)$, can be written in the form

$$\sigma_i(A) = \sup_{\dim(X)=i} \inf_{\substack{x \in X \\ \|x\|_2 = 1}} \|Ax\|_2, \quad (i = 1,2,\ldots,n) , \tag{2.3}$$

where

$$\|y\|_2 = \sqrt{y^H y} \tag{2.4}$$

is the usual Euclidean norm. This characterization provides a natural convention for numbering the singular values of a rectangular matrix: $A \in \mathbb{C}^{m \times n}$ has n singular values of which n-r are zero; A^H has m singular values of which m-r are zero. The nonzero singular values of

A and A^H are the same.

Two inequalities that we shall need in the sequel follow fairly directly from (2.3). They are

$$\sigma_i(A) - \sigma_1(E) \le \sigma_i(A) \le \sigma_i(A) + \sigma_1(E)$$

and

$$\sigma_i(AC) \le \sigma_i(A)\sigma_1(C),\ \sigma_1(A)\sigma_i(C)\ . \tag{2.5}$$

Unitarily invariant matrix norms. A norm on $\mathbb{C}^{m\times n}$ is a function $\|\cdot\|: \mathbb{C}^{m\times n} \to \mathbb{R}$ that satisfies the conditions

$$\begin{aligned} &1.\quad A \ne 0 \Longrightarrow \|A\| > 0\ ,\\ &2.\quad \|\alpha A\| = |\alpha|\,\|A\|\ ,\\ &3.\quad \|A+B\| \le \|A\| + \|B\|\ . \end{aligned} \tag{2.6}$$

A norm $\|\cdot\|$ is unitarily invariant if

$$\|U^H A V\| = \|A\|$$

for all unitary matrices U and V. The perturbation bounds in this paper will be cast in terms of unitarily invariant norms, whose properties will now be described.

Let U and V be the unitary matrices realizing the singular value decomposition of the matrix $A \in \mathbb{C}^{m\times n}$. Then for any unitarily invariant norm $\|\cdot\|_{m,n}$

$$\|A\|_{m,n} = \|U^H AV\|_{m,n} = \left\| \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \right\|_{m,n}.$$

Thus $\|A\|_{m,n}$ is a function of the singular values of A, say

$$\text{(2.7)} \qquad \|A\|_{m,n} = \varphi_{m,n}(\sigma_1,\sigma_2,\ldots,\sigma_n) .$$

It follows from (2.6) that $\varphi_{m,n}$ regarded as a function on $\mathbb{R}^n$ is a norm. Since the interchange of two rows or two columns of a matrix is a unitary transformation of the matrix, the function $\varphi_{m,n}$ is symmetric in its arguments $\sigma_1,\sigma_2,\ldots,\sigma_n$. It can also be shown that $\varphi_{m,n}$ is nondecreasing in the sense that

$$\text{(2.9)} \qquad 0 \le \sigma_i \le \sigma_i' \quad (i=1,2,\ldots,n) \Rightarrow \varphi_{m,n}(\sigma_1,\ldots,\sigma_n) \le \varphi_{m,n}(\sigma_1',\ldots,\sigma_n') .$$

We shall say that the norm $\|\cdot\|_{m,n}$ is <u>generated</u> by $\varphi_{m,n}$.

An important norm is the spectral norm $\|\cdot\|_2$ generated by the function φ defined by

$$\varphi(\sigma_1,\sigma_2,\ldots,\sigma_n) = \max\{|\sigma_1|,\ldots,|\sigma_n|\} .$$

This norm can also be defined by the equation

$$\text{(2.10)} \qquad \|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2 ,$$

where $\|\cdot\|_2$ on the right denotes the Euclidean norm defined by (2.4).

The spectral norm satisfies an important consistency relation with other unitarily invariant norms. If $\|\cdot\|$ is a unitarily invariant norm generated by φ, then it follows from (2.5) and (2.9) that

$$\|CD\| \le \|C\|_2\|D\|, \ \|C\|\|D\|_2 \tag{2.11}$$

whenever the product CD is defined. In particular $\|\cdot\|_2$ is consistent with itself over matrices and vectors of all dimensions.

A second example of a unitarily invariant norm is the Frobenius norm generated by the function

$$\varphi_F(\sigma_1,\sigma_2,\ldots,\sigma_n) = (\sigma_1^2+\sigma_2^2+\ldots+\sigma_n^2)^{1/2} .$$

For any matrix $A \in \mathbb{C}^{m\times n}$

$$\|A\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} |\alpha_{ij}^2| = \operatorname{trace}(A^H A) .$$

The Frobenius norm satisfies the consistency relation

$$\|CD\|_F \le \|C\|_F\|D\|_F .$$

Since we shall be dealing with matrices of varying dimensions, we shall work with a family of unitarily invariant norms defined on $\bigcup_{m,n=1}^{\infty} \mathbb{C}^{m\times n}$. It is important that the individual norms so defined interact with one another properly. Accordingly, we make the following definition.

<u>Definition</u> 2.1. Let $\|\cdot\|: \bigcup_{m,n=1}^{\infty} \mathbb{C}^{m\times n} \to \mathbb{R}$ be a family of unitarily invariant norms. Then $\|\cdot\|$ is <u>uniformly generated</u> if there is a symmetric function φ, defined for all infinite sequences with only a finite number of nonzero terms, such that

$$\|A\| = \varphi(\sigma_1(A), \sigma_2(A), \ldots, \sigma_n(A), 0, 0, \ldots)$$

for all $A \in \mathbb{C}^{m\times n}$. It is <u>normalized</u> if

$$\|x\| = \|x\|_2$$

for any vector x.

The function φ in the above definition must satisfy the conditions (2.6). Any norm defined by such a function can be normalized. Indeed we have

$$\|x\| = \varphi(\sigma_1(x), 0, 0, \ldots) = \varphi(\|x\|_2, 0, 0, \ldots) ,$$

from which it follows that $\|x\| = \mu\|x\|_2$ for some constant μ that is independent of the dimension of x. The function $\mu^{-1}\varphi$ then generates the normalized family of norms.

A uniformly generated family of norms has some nice properties. First, since the nonzero singular values of a matrix and its conjugate transpose are the same, we have

$$\|A^H\| = \|A\| .$$

Second, if a matrix is bordered by zero matrices, its norm remains unchanged; i.e.,

$$\text{(2.12)} \qquad \|A\| = \left\| \begin{pmatrix} 0 & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\| .$$

In particular if A is in reduced form, then

$$\|A\| = \|A_{11}\|$$

and

$$\|A^{\dagger}\| = \|A_{11}^{-1}\| .$$

A third property is that if $\|\cdot\|$ is normalized then

$$\|A\|_2 \le \|A\| . \tag{2.13}$$

In fact from (2.11) and the fact that $\|x\| = \|x\|_2$, we have

$$\|Ax\|_2 = \|Ax\| \le \|A\|\|x\|_2 \tag{2.14}$$

for all x. But by (2.10) $\|A\|_2$ is the smallest number for which (2.14) holds, from which (2.13) follows. A trivial corollary of (2.11) and (2.14) is that $\|\cdot\|$ is consistent:

$$\|CD\| \le \|C\|\|D\| .$$

Finally we observe that

$$\forall x \ \|Cx\|_2 \le \|Dx\|_2 \Longrightarrow \|C\| \le \|D\|. \tag{2.15}$$

To prove this implication note that by (2.3) the hypothesis implies that $\sigma_i(C) \le \sigma_i(D)$. Hence the inequality $\|C\| \le \|D\|$ follows from (2.9).

In the sequel $\|\cdot\|$ will always refer to a uniformly generated, normalized, unitarily invariant norm.

Perturbation of matrix inverses. We shall later need some results on the inverses of perturbations of nonsingular matrices. These are summarized in the following theorem.

Theorem 2.2. Let A be nonsingular and suppose that

$$\|A^{-1}\|_2\|E\| < 1 .$$

Then $A + E$ is nonsingular,

$$\|(A+E)^{-1}\| \le \frac{\|A^{-1}\|}{\gamma} ,$$

and

$$\frac{\|(A+E)^{-1}-A^{-1}\|}{\|A^{-1}\|} \le \frac{\kappa}{\gamma}\,\frac{\|E\|}{\|A\|} , \tag{2.16}$$

where

$$\kappa = \|A\|\|A^{-1}\|_2 \tag{2.17}$$

and

$$\gamma = 1 - \kappa\frac{\|E\|}{\|A\|} > 0 .$$

In most applications of Theorem 2.2 the number γ is insignificantly different from unity. The number κ is usually called the condition number of A (with respect to inversion). It measures the sensitivity of the inverse of A to perturbations in A. Similarly defined quantities will play similar roles in the perturbation theory for the pseudo-inverse.

Projections. We have already observed that the orthogonal projections P_A and R_A onto the column space and the row space of A can be expressed in terms of the pseudo-inverse. The projection onto $R(A)^{\perp}$ will be denoted by

$$P_A^{\perp} \equiv I - P_A .$$

Likewise

$$R_A^{\perp} \equiv I - R_A$$

will denote the projection onto $R(A^H)^{\perp}$.

When A is in reduced form, its projections can be easily written out:

$$P_A = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{C}^{m\times m} ,$$

$$R_A = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{C}^{n\times n} .$$

It follows that

$$\|P_A A R_A\| = \|A_{11}\|$$

and

$$\|P_A E R_A\| = \|E_{11}\| , \quad \|P_A E R_A^{\perp}\| = \|E_{12}\|$$

$$\|P_A^{\perp} E R_A\| = \|E_{21}\| , \quad \|P_A^{\perp} E R_A^{\perp}\| = \|E_{22}\| .$$

These identities enable us to pass from results for the reduced form to general results stated in terms of projections of A and E.

We shall need some properties of norms of projections later. These are summarized in the following theorem.

Theorem 2.3. For any A and B the following statements are true.

1. If rank (A) = rank (B), then the singular values of $P_A P_B^{\perp}$ and $P_B P_A^{\perp}$ are the same so that

$$\|P_A P_B^{\perp}\| = \|P_B P_A^{\perp}\| .$$

Moreover the nonzero singular values σ^2 of $P_A P_B^{\perp}$ correspond to pairs $\pm\sigma$ of eigenvalues of $P_B - P_A$, so that

$$\|P_B - P_A\|_2 = \|P_A P_B^{\perp}\|_2 = \|P_B P_A^{\perp}\|_2 .$$

2. If $\|P_B - P_A\|_2 < 1$, then rank (A) = rank (B).
3. If rank (B) $\geq$ rank (A), then

$$\|P_B P_A^{\perp}\| \geq \|P_B^{\perp} P_A\| .$$

Proof. Proofs of parts one and two are readily found in the literature. For part three write $P_B = P_1 + P_2$ where rank (P_1) = rank (A) and $P_A P_2 = 0$ (i.e. $R(P_2)$ is orthogonal to $R(A)$). Then

$$\|P_A P_B^{\perp}\| = \|P_A(I - P_1 - P_2)\| = \|P_A(I - P_1)\| = \|P_1 P_A^{\perp}\| ,$$

the last equality following from part 1. Now for any x

$$\|P_1 P_A^\perp x\| \le \|P_B P_A^\perp x\| ,$$

and the result follows from (2.15). □

When $B = A + E$, we can estimate $\|P_B P_A^\perp\|$ in terms of E.

Theorem 2.4. The product $P_B P_A^\perp$ can be written in the form

$$P_B P_A^\perp = (B^\dagger)^H R_B E^H P_A^\perp . \tag{2.18}$$

Hence

$$\|P_B P_A^\perp\| \le \|B^\dagger\|_2 \|E\| , \tag{2.19}$$

and if rank (A) = rank (B), then

$$\|P_B P_A^\perp\| \le \min \{\|B^\dagger\|_2, \|A^\dagger\|_2\} \; \|E\| . \tag{2.20}$$

Proof. We have

$$\begin{aligned} P_B P_A^\perp = P_B^H P_A^\perp &= (B^\dagger)^H B^H P_A^\perp \\ &= (B^\dagger)^H (A+E)^H P_A^\perp = (B^\dagger)^H E^H P_A^\perp \\ &= (B^\dagger)^H B^H (B^\dagger)^H E^H P_A^\perp = (B^\dagger)^H R_B E^H P_A^\perp , \end{aligned}$$

which establishes (2.18). The inequality (2.19) follows upon taking norms in (2.18). Finally (2.20) follows from part 1 of Theorem 2.3. □

Theorems 2.3 and 2.4 have obvious analogues for other combinations of projectors (e.g. $R_B^\perp R_A = -A^\dagger E R_B^\perp$). In the sequel a reference to these theorems will also cover any trivial variants.

The case when $\|P_B-P_A\|_2 < 1$ will be particularly important later. We have seen in part 2 of Theorem 2.3 that in this case rank (A) = rank (B). However more is true: no vector in $R(A)$ can be orthogonal to $R(B)$ and vice versa. For suppose that $x \neq 0$ satisfies $P_A x = x$ and $P_B x = 0$. Then $(P_B-P_A)x = -x$, which implies that $\|P_B-P_A\|_2 \geq 1$. Conversely if $\|P_B-P_A\|_2 = 1$ then there is a vector in $R(A)$ or $R(B)$ that is orthogonal to $R(B)$ or $R(A)$. To see this, note that by Theorem 2.3.1 there is a vector x such that $(P_B-P_A)x = x$. If $P_A x = 0$ then $P_B x = x$ which shows that $x \in R(B)$ and $x \in R(A)^{\perp}$. If, on the other hand, $P_A x \neq 0$, then since $P_A x = -(I-P_B)x$ we have $P_B(P_A x) = 0$, which shows that $P_A x \in R(A)$ and $P_A x \in R(B)^{\perp}$.

Because of the above considerations, we shall say that $R(A)$ and $R(B)$ are <u>acute</u> whenever $\|P_B-P_A\|_2 < 1$. The following theorem gives sufficient conditions for $R(A)$ and $R(B)$ to be acute.

<u>Theorem</u> 2.5. If rank (A) = rank (B) and

$$\|A^+\|_2 \|P_A E R_A\|_2 < 1 ,$$

then $R(A)$ and $R(B)$ are acute.

<u>Proof</u>. We shall use the reduced form. From Theorem 2.2 it follows that $B_{11} = A_{11} + E_{11}$ nonsingular. Hence

$$\text{rank} \left[\begin{pmatrix} B_{11} \\ E_{21} \end{pmatrix} \right] = \text{rank (A)} = \text{rank (B)} .$$

It follows that

$$R(B) = R\left[\begin{pmatrix} B_{11} \\ E_{21} \end{pmatrix}\right] = R\left[\begin{pmatrix} I_r \\ E_{21}B_{11}^{-1} \end{pmatrix}\right].$$

But

$$R(A) = R\left[\begin{pmatrix} I_r \\ 0 \end{pmatrix}\right],$$

from which it is easily seen that no vector in $R(A)$ can be orthogonal to $R(B)$ and vice versa. □

Theorem 2.4 shows that if rank (A) = rank (B) the spaces $R(A)$ and $R(B)$ are acute whenever E is sufficiently small. For this reason we shall say that B is an acute perturbation of A if A and B satisfy the hypotheses of Theorem 2.4. The reader should remember that the statement "B is an acute perturbation of A" is stronger than the statement "$R(B)$ and $R(A)$ are acute."

Notes and References. The properties of singular values are well known. See Stewart (1973) for an introduction and Gohberg and Krein (1965) for a more detailed treatment in an infinite dimensional setting.

Von Neumann (1937) was the first to prove that unitarily invariant norms can be written as a function of singular values (the function $\varphi_{m,n}$ in (2.7) is usually called a symmetric gauge function). Systematic treatments of unitarily invariant norms may be found in Mirsky (1960) and Gohberg and Krein (1965).

The treatment of unitarily invariant norms in finite dimensional spaces has often been a little sloppy. In infinite dimensional settings there is usually only one space and one generating function, and the same is true in a finite dimensional setting when one is concerned with square matrices. However, when one considers rectangular matrices with varying dimensions, different norms can be used for different dimensions, and there is no reason why these norms should interact nicely. How bad things can get is illustrated by the family or norms $\|\ \|$ defined for $A \in \mathbb{C}^{m\times n}$ by

$$\|A\| = \frac{m}{n}\|A\|_2 .$$

This family is unitarily invariant and consistent, but $\|A^H\| \neq \|A\|$, unless A is square, and the relation (2.15) does not hold in general. Definition 2.1 represents a return to the simplicity of the infinite dimensional case.

Theorem 2.2 is classical and is usually proved by an appeal to the Neumann series representation $(I-A)^{-1} = I + A + A^2 + \ldots$. Wilkinson (1965) gives a proof that does not use series and discusses at some length the notion of condition number. The result is usually proved under the assumption that $\|I\| = 1$; however, the proofs can be extended to establish the result for any consistent norm.

The results in Theorem 2.3 are well known to people who work closely with orthogonal projectors (for proofs see Afriat (1957) or Wedin (1969)). The decomposition in Theorem 2.3 was established in a slightly weaker form by Wedin (1973). In some cases, when E is small, R_B will be near R_A

and the approximation $\|P_A E R_B\| \cong \|E_{21}\|$ will be more realistic in (2.19).

The number $\|P_B - P_A\|_2$ is closely related to various measures of separation between subspaces. See Kato (1966) and especially Davis and Kahan (1970) where further references may be found. Theorem 2.4, with $\|P_A E R_A\|$ replaced by $\|E\|$, is proved by Wedin (1973). The term "acute" ordinarily refers to the angle subtended by two line segments, not to the segments themselves, and it is technically misapplied when subspaces are said to the acute. But this usage will cause no confusion and it is better than the ugly phrase "in the acute case." The term "acute perturbation" is new.

3. The Pseudo-Inverse

In this section we shall consider the problem of bounding $\|B^\dagger - A^\dagger\|$ in terms of $\|E\|$. We shall obtain three basic theorems: one for when rank (A) $\neq$ rank (B), one for when rank (A) = rank (B), and one for when B is an acute perturbation of A. All these theorems are based on expressions for $B^\dagger$, which also yield asymptotic expressions for $B^\dagger$ and expressions for the derivative of $A^\dagger$.

Lower bounds. Before proceeding to obtain bounds on $\|B^\dagger - A^\dagger\|$, we shall show how bad things can be by deriving lower bounds.

Theorem 3.1. If $\mathcal{R}(A)$ and $\mathcal{R}(B)$ are not acute, then

$$\|B^\dagger - A^\dagger\|_2 \geq 1/\|E\|_2 . \tag{3.1}$$

If further rank (B) $\geq$ rank (A), then

$$(3.2) \qquad \|B^{\dagger}\|_2 \geq 1/\|E\|_2 .$$

<u>Proof</u>. Suppose for definiteness that rank (B) $\geq$ rank (A). Then there is a vector $y \in R(B)$ with $\|y\|_2 = 1$ such that $y \in R(A)^{\perp}$. Thus

$$1 = y^H y = y^H P_B y = y^H BB^{\dagger} y = y^H (A+E)B^{\dagger} y$$
$$= y^H EB^{\dagger} y \leq \|E\|_2 \|B^{\dagger} y\|_2 ,$$

which shows that $\|B^{\dagger}y\|_2$, and hence $\|B^{\dagger}\|_2$ is not less than $1/\|E\|_2$. From this and the fact that $A^{\dagger}y = A^{\dagger}P_A y = 0$ we have

$$\frac{1}{\|E\|_2} \leq \|B^{\dagger}y\|_2 \leq \|(B^{\dagger}-A^{\dagger})y\|_2 \leq \|B^{\dagger}-A^{\dagger}\|. \ \square$$

Theorem 3.1 shows that the pseudo-inverse of a general matrix is not a continuous function of its elements, unless the class of perturbations is restricted. It also says that if two nearby matrices do not have acute column spaces, then one of them at least must have a large pseudo-inverse. Moreover if they are of the same rank, then both of them must have large pseudo-inverses.

<u>A decomposition of</u> $B^{\dagger} - A^{\dagger}$. In spite of the negative results in Theorem 3.1, it is possible to obtain bounds on $\|B^{\dagger}-A^{\dagger}\|$ in the general case, although these bounds need not remain finite as B approaches A. The basis for obtaining such bounds is contained in the following theorem.

<u>Theorem</u> 3.2. The following two decompositions of $B^{\dagger} - A^{\dagger}$ are valid:

$$B^{\dagger} - A^{\dagger} = -B^{\dagger}P_B E R_A A^{\dagger} + B^{\dagger}P_B P_A^{\perp} - R_B^{\perp} R_A A^{\dagger} \ , \tag{3.3}$$

$$B^{\dagger} - A^{\dagger} = -B^{\dagger}P_B E R_A A^{\dagger} + (B^H B)^{\dagger} R_B E^H P_A^{\perp} - R_B^{\perp} E^H P_A (AA^H)^{\dagger} \ . \tag{3.4}$$

Proof. Both expressions can be verified directly by replacing E with $B - A$, replacing the projectors by their expressions in terms of pseudo-inverses, and simplifying. □

It should be noted that (3.3) can be obtained directly from (3.2) by using Theorem 2.4 to express $P_B P_A^{\perp}$ and $R_B^{\perp} R_A$ in terms of E.

The general theorem. We are now in a position to prove the general theorem bounding $\|B^{\dagger}-A^{\dagger}\|$.

Theorem 3.3. For any A and B with $B = A + E$,

$$\|B^{\dagger}-A^{\dagger}\| \leq \mu \max \{\|A^{\dagger}\|_2^2, \|B^{\dagger}\|_2^2\} \|E\| \ ,$$

where μ is given in the following table:

$\|\cdot\|$	arbitrary	spectral	Frobenius
μ	3	$\frac{1+\sqrt{5}}{2}$	$\sqrt{2}$

Proof. The proof is a slight modification of the proof given by Wedin (1973). We shall give only the proof for the Frobenius norm.

Suppose for definiteness rank (B) ≤ rank (A). Let F_1, F_2, and F_3 denote the three terms on the right-hand side of (3.3). Then the column spaces of F_1 and F_2 are orthogonal to the column space of F_3. Hence

$$\|B^{\dagger}-A^{\dagger}\|_F^2 = \|F_1+F_2\|_F^2 + \|F_3\|_2^2 . \tag{3.5}$$

Now since $F_1 + F_2 = B^{\dagger}(P_B DA^{\dagger}P_A + P_B P_A^{\perp})$,

$$\|F_1+F_2\|_F^2 \le \|B^{\dagger}\|_2^2(\|P_B EA^{\dagger}P_A\|_F^2 + \|P_B P_A^{\perp}\|_F^2) .$$

But from Theorems 2.4 and 2.5

$$\begin{aligned} &\|P_B EA^{\dagger}P_A\|_F^2 + \|P_B P_A^{\perp}\|_F^2 \\ &\le \|P_B EA^{\dagger}\|_F^2 + \|P_B^{\perp}P_A\|_F^2 \\ &= \|P_B EA^{\dagger}\|_F^2 + \|P_B^{\perp}EA^{\dagger}\|_F^2 \\ &= \|EA^{\dagger}\|_F^2 \le \|E\|_F^2\|A^{\dagger}\|_2^2 . \end{aligned}$$

Hence

$$\|F_1+F_2\|_F \le \|A^{\dagger}\|_2\|B^{\dagger}\|_2\|E\|_F . \tag{3.6}$$

Also from Theorem 2.5

$$\begin{aligned} \|F_3\|_F &= \|A^{\dagger}\|_2\|R_B^{\perp}R_A\|_F = \|A^{\dagger}\|_2\|R_A R_B^{\perp}\|_F \\ &= \|A^{\dagger}\|_2\|A^{\dagger}ER_B^{\perp}\|_F \le \|A^{\dagger}\|_2^2\|E\|_F , \end{aligned} \tag{3.7}$$

and the result follows on combining (3.3), (3.6), and (3.7). Since the final bound is symmetric in A and B, it also holds when rank $(B) \ge$ rank (A). □

It should be noted that these bounds do not imply that $\|B^{\dagger}-A^{\dagger}\|$ is small when $\|E\|$ is small, since $B^{\dagger}$ may grow unboundedly as E approaches zero.

The case rank (A) = rank (B). When A and B have the same rank, we can strengthen Theorem 3.3 in two ways. First, we can replace the term $\max\{\|A^{\dagger}\|_2^2, \|B^{\dagger}\|_2^2\}$ with the product $\|A\|_2\|B\|_2$. Second we can distinguish more cases for the constant μ. In the following theorem recall that $A \in \mathbb{C}^{m\times n}$ with $m \geq n$.

Theorem 3.4. If rank (A) = rank (B), then

$$\|B^{\dagger}-A^{\dagger}\| \leq \mu\|A^{\dagger}\|_2\|B^{\dagger}\|_2\|E\| , \tag{3.8}$$

where μ is given in the following table.

rank \ $\|\cdot\|$	Arbitrary	Spectral	Frobenius
rank (A) < min(m,n)	3	$(1+\sqrt{5})/2$	$\sqrt{2}$
rank (A) = min(m,n), $m \neq n$	2	$\sqrt{2}$	1
rank (A) = m = n	1	1	1

The proof of this theorem may be found in Wedin (1973). The bound (3.8) may be recast in the form

$$\frac{\|B^{\dagger}-A^{\dagger}\|}{\|B^{\dagger}\|_2} \leq \mu\kappa\frac{\|E\|}{\|A\|} , \tag{3.9}$$

where

$$\kappa = \|A\|\|A^{\dagger}\|_2 .$$

In this form the result is almost analogous to the bound (2.16) for the inverse in Theorem 2.2. The bound (3.9) also implies that as E approaches zero, the relative error in $B^\dagger$ approaches zero, which further implies that $B^\dagger$ approaches $A^\dagger$. Remembering, on the other hand, that if rank $(B) \neq$ rank (A) then $\mathcal{R}(A)$ and $\mathcal{R}(B)$ cannot be acute, we have from Theorem 3.1 the following corollary of Theorem 3.4.

Corollary 3.5. A necessary and sufficient condition that

$$\lim_{B\to A} B^\dagger = A^\dagger$$

is that rank (B) = rank (A) as B approaches A.

Acute Perturbations. It is evident from the proofs of Theorems 3.3 and 3.4 that we have given away much in deriving the bounds. In particular, *if B is a small acute perturbation of A then P_A and P_B are nearly* equal, and the same is true of R_A and R_B. Thus it follows from (3.4) that $B^\dagger - A^\dagger$ can be decomposed into three terms--one essentially depending on $P_A E R_A$, one on $P_A E R_A^\perp$, and one on $P_A^\perp E R_A^\perp$. However, this does not tell the whole story; for we shall show that the dependency of $B^\dagger - A^\dagger$ on $P_A E R_A^\perp$ and $P_A^\perp E R_A$ is bounded, *no matter how large these projections may be*.

In order to state our theorems concisely, we must first introduce some additional notation. Let $\|\cdot\|$ be generated by φ and for any $F \in \mathbb{C}^{k\times r}$ $(k \geq r)$ define

$$\psi_\varphi(F) = \varphi\left[\frac{\sigma_1(F)}{[1+\sigma_1^2(F)]^{1/2}}, \ldots, \frac{\sigma_r(F)}{[1+\sigma_r^2(F)]^{1/2}}\right] \quad . \tag{3.10}$$

The function ψ_φ is not a norm; however, it has some useful properties. First, from (2.5) and the monotonicity of φ,

$$\psi(GF) \le \psi(\|G\|_2 F) \le \psi(\|G\|F) .$$

Second, since for $\alpha \ge 1$

$$\frac{\alpha\sigma}{(1+\alpha\sigma^2)^{1/2}} \le \frac{\alpha\sigma}{(1+\sigma^2)^{1/2}} ,$$

we have

$$\alpha \ge 1 \Longrightarrow \psi_\varphi(\alpha F) \le \alpha\psi_\varphi(F) .$$

For small F, $\psi_\varphi(F)$ is asymptotic to $\|F\|$:

$$\psi_\varphi(F) = \|F\| + o(\|F\|) .$$

For large F, ψ_φ is bounded:

$$\psi_\varphi(F) \le \|I_r\| .$$

Finally, for the spectral norm

$$\psi_2(F) = \frac{\|F\|_2}{(1+\|F\|_2^2)^{1/2}} .$$

Our first result concerns a rather special matrix.

<u>Lemma</u> 3.6. The matrix

$$\begin{pmatrix} I \\ F \end{pmatrix}$$

satisfies

$$\left\| \begin{pmatrix} I \\ F \end{pmatrix}^{\dagger} \right\|_2 \le 1 \tag{3.11}$$

and

$$\left\| \begin{pmatrix} I \\ F \end{pmatrix}^{\dagger} - (I \quad 0) \right\| = \psi_{\varphi}(F) \ . \tag{3.12}$$

<u>Proof</u>. It is easily verified that

$$\begin{pmatrix} I \\ F \end{pmatrix}^{\dagger} = (I+F^H F)^{-1}(I \quad F^H) \ , \tag{3.13}$$

whose singular values are

$$\frac{1}{[1+\sigma_i^2(F)]^2} \le 1 \ ,$$

from which (3.11) follows. Also if

$$G = \begin{pmatrix} I \\ F \end{pmatrix}^{\dagger} - (I \quad 0) \ ,$$

then

$$GG^H = I - (I+F^H F)^{-1} \ .$$

It follows that the singular values of G are given by

$$\frac{\sigma_i(F)}{[1+\sigma_i(F)]^{1/2}} \ ,$$

which establishes (3.12). □

The main result is based on an explicit representation of $B^{\dagger}$. We shall work with the reduced forms of A and B.

<u>Theorem</u> 3.7. Let B be an acute perturbation of A. Then

$$B^{\dagger} = (I \quad F_{12})^{\dagger} B_{11}^{-1} \begin{pmatrix} I \\ F_{21} \end{pmatrix}^{\dagger}, \tag{3.14}$$

where

$$F_{21} = E_{21} B_{11}^{-1}, \quad F_{12} = B_{11}^{-1} E_{12}.$$

<u>Proof</u>. As in the proof of Theorem 3.4, we have

$$\mathcal{R}(B) = \mathcal{R}\left[\begin{pmatrix} B_{11} \\ E_{21} \end{pmatrix}\right].$$

Thus the columns of

$$\begin{pmatrix} E_{12} \\ E_{22} \end{pmatrix}$$

can be expressed as a linear combination of the columns of

$$\begin{pmatrix} B_{11} \\ E_{21} \end{pmatrix}.$$

Since $B_{11}(B_{11}^{-1}E_{12}) = E_{12}$, we must have

$$\begin{pmatrix} E_{12} \\ E_{22} \end{pmatrix} = \begin{pmatrix} B_{11} \\ E_{21} \end{pmatrix} B_{11}^{-1} E_{12},$$

from which it follows that

$$B = \begin{pmatrix} I \\ F_{21} \end{pmatrix} B_{11}^{-1} (I \quad F_{12}) . \tag{3.15}$$

The result now follows from Penrose's conditions. □

It is interesting to observe that, from (3.15),

$$B_{22} = E_{22} = F_{21} B_{11}^{-1} F_{12} ,$$

which is of second order in $\|E\|$. In other words, if rank (A+E) = rank (A), then $P_A^{\perp} E R_A^{\perp}$ must approach zero quadratically as E approaches zero.

We turn now to the perturbation theorem.

<u>Theorem</u> 3.8. Let B be an acute perturbation of A. Let

$$\kappa = \|A\| \|A^{\dagger}\|_2$$

and let

$$\gamma = 1 - \kappa \frac{\|E_{11}\|}{\|A\|} > 0 .$$

Then

$$\|B^{\dagger}\| \le \frac{\|A^{\dagger}\|}{\gamma} ,$$

and

$$\frac{\|B^{\dagger} - A^{\dagger}\|}{\|A^{\dagger}\|} \le \frac{\kappa}{\gamma} \frac{\|E_{11}\|}{\|A\|} + \psi_{\varphi}\left(\frac{\kappa}{\gamma} \frac{\|E_{12}\|}{\|A\|}\right) + \psi_{\varphi}\left(\frac{\kappa}{\gamma} \frac{\|E_{12}\|}{\|A\|}\right) , \tag{3.16}$$

where ψ_{φ} is defined by (3.10).

<u>Proof</u>. Let F_{ij} be defined as in Theorem 3.7. Let

$$I_{21} = \begin{pmatrix} I_r \\ 0 \end{pmatrix}, \quad I_{12} = (I_r \quad 0)$$

$$J_{21} = \begin{pmatrix} I_r \\ F_{21} \end{pmatrix}, \quad J_{12} = (I_r \quad F_{12}) .$$

It follows from Theorem 2.2 that

$$\|B_{11}^{-1}\| = \|(A_{11}+E_{11})^{-1}\| \le \frac{\|A_{11}^{-1}\|}{\gamma} = \frac{\|A^{\dagger}\|}{\gamma} .$$

From (3.14), $B^{\dagger} = J_{12}^{\dagger}B_{11}^{-1}J_{21}^{\dagger}$, and from Lemma 3.6

$$\|B^{\dagger}\| \le \|J_{12}^{\dagger}\|_2\|B_{11}^{-1}\|\|J_{21}^{\dagger}\|_2 \le \|B_{11}^{-1}\| \le \frac{\|A^{\dagger}\|}{\gamma} .$$

Now from (3.14)

$$B^{\dagger}-A^{\dagger} = (J_{12}^{\dagger}-I_{12}^{\dagger})A_{11}^{-1}I_{21}^{\dagger} + J_{12}^{\dagger}A_{11}^{-1}(J_{21}^{\dagger}-I_{21}^{\dagger}) + J_{12}^{\dagger}(B_{11}^{-1}-A_{11}^{-1})J_{21}^{\dagger} . \tag{3.17}$$

From Theorem 2.2 we have the following bound:

$$\|J_{12}^{\dagger}(B_{11}^{-1}-A_{11}^{-1})\| \le \|A_{11}^{-1}\|_2 \frac{\kappa}{\gamma} \frac{\|E_{11}\|}{\|A_{11}\|} . \tag{3.18}$$

By Lemma 3.6

$$\begin{aligned} \|(J_{12}^{\dagger}-I_{12}^{\dagger})A_{11}^{-1}I_{21}^{\dagger}\| &\le \|A_{11}^{-1}\|_2\|J_{12}^{\dagger}-I_{12}^{\dagger}\| \\ &= \|A_{11}^{-1}\|_2\psi_{\varphi}(F_{12}) \\ &= \|A_{11}^{-1}\|_2\psi_{\varphi}(B_{11}^{-1}E_{12}) \\ &\le \|A_{11}^{-1}\|_2\psi_{\varphi}\left(\frac{\kappa}{\gamma} \frac{\|E_{12}\|}{\|A\|}\right) , \end{aligned} \tag{3.19}$$

and likewise

$$(3.20)\qquad \|J_{12}^{\dagger}A_{11}^{-1}(J_{21}^{\dagger}-I_{21}^{\dagger})\| \le \|A_{11}^{-1}\|_2\psi_\varphi\left(\frac{\kappa}{\gamma}\frac{\|E_{21}\|}{\|A\|}\right) .$$

The bound (3.16) follows on combining (3.17), (3.18), (3.19), and (3.20) and remembering that $\|A_{11}^{-1}\| = \|A^{\dagger}\|$. □

The bound (3.16) gives a rather nice dissection of $\|B^{\dagger}-A^{\dagger}\|$. Asymptotically, for E small, it reduces to the bound that would be obtained by taking norms in (3.4); i.e.,

$$\frac{\|B^{\dagger}-A^{\dagger}\|}{\|A^{\dagger}\|} \le \frac{\kappa}{\gamma}\,\frac{\|E_{11}\|+\|E_{12}\|+\|E_{21}\|}{\|A\|} .$$

However, the bound additionally shows that E_{12} and E_{21} can have at most a bounded effect on $\|B^{\dagger}-A^{\dagger}\|$.

When A is square and nonsingular, E_{12} and E_{21} are void, and the bound reduces to that of Theorem 2.2. Note that the number κ, defined in analogy with (2.17), plays an analogous role here.

<u>Asymptotic forms and derivatives</u>. Asymptotic forms for B may be obtained from either (3.4) or (3.14). Of course for $B^{\dagger}$ to approach $A^{\dagger}$ we must have rank (A) = rank (B); and since we are assuming that E is arbitrarily small, by Theorem 2.5 we have that B is an acute perturbation of A. In this case

$$B^{\dagger} = A^{\dagger} + O(\|E\|) ,$$

and

$$P_B = BB^{\dagger} = (A+E)[A^{\dagger}+O(\|E\|)] = P_A + O(\|E\|)$$

with similar expressions for the other projections. Hence from (3.4)

$$B^{\dagger} = A^{\dagger} - A^{\dagger}P_A E R_A A^{\dagger} + (A^H A)^{\dagger} R_A E^H P_A^{\perp}$$

(3.21)

$$- R_A^{\perp} E^H P_A (AA^H)^{\dagger} + O(\|E\|^2) .$$

It follows immediately from (3.21) that if $A(\tau)$ is a differentiable function of τ with

$$\text{rank } [A(\tau)] = \text{rank } [A(\tau')]$$

for all τ, then $A(\tau)^{\dagger}$ is a differentiable function of τ and

$$\frac{dA^{\dagger}}{d\tau} = -A^{\dagger}P_A \frac{dA}{d\tau} R_A A^{\dagger} + (A^H A)^{\dagger} R_A \frac{dA^H}{d\tau} P_A^{\perp}$$

(3.22)

$$- R_A^{\perp} \frac{dA^H}{d\tau} P_A (AA^H)^{\dagger} .$$

The asymptotic form obtained from (3.14) can be useful computationally when A has been put in reduced form as a preliminary to computing $A^{\dagger}$. We have from (3.21) that

$$B_{11}^{-1} = A_{11}^{-1} - A_{11}^{-1}E_{11}A_{11}^{-1} + O(\|E_{11}\|^2) .$$

From (3.13) in the proof of Lemma 3.6 we have

$$\begin{pmatrix} I \\ F_{21} \end{pmatrix}^{\dagger} = (I \;\; A_{11}^{-H}E_{21}^H) + O(\|E_{11}\|\|E_{21}\|)$$

and

$$(I \quad F_{12})^{\dagger} = \begin{pmatrix} I \\ E_{12}^{H} A_{11}^{-H} \end{pmatrix} + O(\|E_{11}\| \|E_{12}\|) .$$

Hence from (3.14)

$$(3.23) \qquad B^{\dagger} = \begin{pmatrix} A_{11}^{-1} - A_{11}^{-1}E_{11}A_{11}^{-1} + O(\|E_{11}\|^2) & (A_{11}^{H}A_{11})^{-1}E_{21}^{H} + O(\|E_{11}\|\|E_{21}\|) \\ E_{12}^{H}(A_{11}A_{11}^{H})^{-1} + O(\|E_{12}\|\|E_{12}\|) & E_{12}^{H}(A_{11}^{H}A_{11}A_{11}^{H})^{-1}E_{21}^{H} \\ & + O(\|E_{11}\|^2\|E_{12}\|\|E_{21}\|) \end{pmatrix} .$$

This expression is in perfect agreement with (3.21) when the E_{ij} are interpreted appropriately as projections of E.

<u>Notes and references</u>. For expository reasons the results of this section have not been presented in the historical order of their development. Penrose (1955) established Corollary 3.4 using techniques that do not give explicit perturbation bounds. The subject was revived by Golub and Wilkinson (1966), whose interest in stable algorithms for solving least squares problems [cf. Golub (1965)] led them to derive first-order perturbation bounds for least squares solutions (more of this later). The first perturbation bounds for the pseudo-inverse itself are given by Ben-Israel (1966), who restricts his class of perturbations so that (in reduced form) only E_{11} is nonzero. More general theorems for acute perturbations were established by Hanson and Lawson (1969), Pereyra (1969), and Stewart (1969). Theorem 3.7 is a refinement and extension of Stewart's bound.

The decompositions (3.3) and (3.4) and the consequent Theorem 3.4 are due to Wedin (1973). Theorem 3.3 is a slight extension of these results. Theorem 3.1 is also due to Wedin (1973), although a slightly restricted form of the result may be found in Stewart (1969). In an earlier report Wedin (1969) considers the sharpness of the constants μ in Theorem 3.4 and shows that for the spectral norm μ cannot be made smaller.

Early differentiability results have been given by Pavel-Parvu and Korganoff (1969) and Hearon and Evans (1968). Wedin (1969) derived the formula (3.22) as we did from the decomposition (3.4). The same result for functions of several variables was derived independently by Golub and Pereyra (1973) in connection with separable nonlinear least squares problems. For further references see Golub and Pereyra (1975).

4. Projections

In this section we shall consider how the projection P_A varies with A. Since $P_A = AA^\dagger$, it might be thought that the perturbation theory for P_A could be derived from the theory developed in the last section for $A^\dagger$. However this approach gives too much away, and sharper bounds may be obtained by working directly with one of the decompositions of $B^\dagger$. In particular we shall work with the decomposition (3.15) based on the reduced forms of A and B.

If $R(A)$ and $R(B)$ are not acute, then $\|P_B - P_A\|_2 = 1$. Consequently we can restrict ourselves to the case where $R(A)$ and $R(B)$ are acute.

More particularly we shall only consider the case where B is an acute perturbation of A.

Theorem 4.1. Let B be an acute perturbation of A, and let κ and γ be defined as in Theorem 3.8. Then

$$\|P_B - P_A\|_2 \le \frac{\frac{\kappa}{\gamma}\,\frac{\|E_{21}\|_2}{\|A\|_2}}{\left[1 + \left(\frac{\kappa}{\gamma}\,\frac{\|E_{21}\|_2}{\|A\|_2}\right)^2\right]^{1/2}} < 1\,. \tag{4.1}$$

Proof. With F_{21} defined as in the last section we have [cf. (3.15)]

$$R(B) = R\begin{pmatrix} I \\ F_{21} \end{pmatrix}.$$

The matrix

$$\begin{pmatrix} I \\ F_{21} \end{pmatrix}(I+F_{21}^H F_{21})^{-1}(I \quad F_{21}^H)$$

is a Hermitian idempotent whose column space is $R(B)$; hence it is P_B. It follows that

$$P_B - P_A = \begin{pmatrix} (I+F_{21}^H F_{21})^{-1} - I & (I+F_{21}^H F_{21})^{-1} F_{21}^H \\ F_{21}(I+F_{21}^H F_{21})^{-1} & F_{21}(I+F_{21}^H F_{21})^{-1} F_{21}^H \end{pmatrix}, \tag{4.2}$$

from which it is easily verified that

$$(4.3) \qquad (P_B - P_A)^2 = \begin{pmatrix} F_{21}^H F_{21}(I + F_{21}^H F_{21})^{-1} & 0 \\ 0 & F_{21}(I + F_{21}^H F_{21})^{-1} F_{21}^H \end{pmatrix}.$$

Now the nonzero singular values of the diagonal blocks in (4.3) are given by

$$\frac{\sigma_i^2(F_{21})}{1 + \sigma_i^2(F_{21})}$$

where the $\sigma_i(F_{21})$ are the nonzero singular values of F_{21}. The result follows from the fact that the largest singular value σ_1 of F_{21} satisfies

$$\sigma_1(F_{21}) = \|F_{21}\|_2 \le \frac{\kappa}{\gamma} \frac{\|E_{21}\|_2}{\|A\|_2} \ . \ \square$$

In terms of projections, the bound (4.1) can be written in the form

$$\|P_B - P_A\|_2 \le \frac{\frac{\kappa}{\gamma} \frac{\|P_A^{\perp} E R_A\|_2}{\|A\|_2}}{\left[1 + \left(\frac{\kappa}{\gamma} \frac{\|P_A^{\perp} E R_A\|_2}{\|A\|_2}\right)^2\right]^{1/2}} \quad 1 .$$

The bound is interesting in several ways. First it depends not at all on E_{12} and E_{22}. Second its dependence on E_{11} is only through the constant γ (in fact the term κ/γ can be replaced throughout by $\|B_{11}^{-1}\|_2 \|A\|_2$). Third the bound is always less than unity. Finally, it goes to zero along with E_{21}. We may summarize this last observation in the following corollary.

<u>Corollary</u> 4.2. Regarding B as variable, a sufficient condition for

$$\lim_{B \to A} P_B = P_A$$

is that rank (A) = rank (B), $\|A^{\dagger}\|_2 \|P_A B R_A\|_2 \le \delta < 1$, and

$$\lim_{B\to A} P_A^{\perp} B R_A = 0 .$$

Asymptotic forms and derivatives. Asymptotic forms may be obtained in the usual way from (4.2). Indeed

$$P_B - P_A = \begin{pmatrix} O(\|E_{21}\|^2) & F_{21}^H + O(\|E_{21}\|^3) \\ F_{21} + O(\|E_{21}\|^3) & O(\|E_{21}\|^2) \end{pmatrix} .$$

In terms of projections

$$P_B = P_A + P_A^{\perp} E R_A A^{\dagger} + A^{\dagger H} R_A E^H P_A^{\perp} + O(\|P_A^{\perp} E R_A\|^2) .$$

It follows that if $A(\tau)$ is differentiable and varies without changing rank, then $P_{A(\tau)}$ is differentiable and

$$\text{(4.4)} \qquad \frac{dP_A}{d\tau} = P_A^{\perp} \frac{dA}{d\tau} R_A A^{\dagger} + A^{\dagger H} R_A \frac{dA^H}{d\tau} P_A .$$

Notes and references. Theorem 4.1 and its corollary appear to be new. The expression (4.4) for the derivative of P_A was first given by Golub and Pereyra (1973).

5. The Linear Least Squares Problem

In this section we shall derive perturbation bounds for the least squares problem of minimizing $\|b-Ax\|_2$. Although the solution of minimum norm is given by $x = A^{\dagger}b$, the perturbation theory of §3 again does not give the best possible results.

We shall assume throughout this section that B is an acute perturbation of A, and we shall work with the reduced form of the problem. In this form x is replaced by $V^H x$ and b is replaced by $U^H b$ (cf. §2). If x and b are partitioned in the forms

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

where $x_1, b_1 \in \mathbb{C}^r$ then

$$x_1 = A_{11}^{-1} b_1 \tag{5.1}$$

and

$$x_2 = 0 .$$

Moreover the norm of the residual vector

$$r = b - Ax$$

is given by

$$\|r\|_2 = \|b_2\|_2 .$$

In the theorems to follow we shall freely use the definitions made in the previous sections [e.g., κ and γ]. One additional piece of notation will be needed; namely, we shall define η as that nonnegative constant such that

$$\|b_1\|_2 = \eta \|A\|_2 \|x\|_2 .$$

Since $b_1 = A_{11}x_1$, we have $\eta \le 1$. Also $\|x\| \le \|A^\dagger\|\|b_1\|$, which shows that $\eta \ge \kappa^{-1}$. When A is ill-conditioned, that is when $A^\dagger$ is large, the vector x may be either large or small. In the first case η is near zero, and we shall say that "x reflects the ill-condition of A."

We first consider perturbations in the vector b.

Theorem 5.1. Let $x = A^\dagger b$ and $x + h = A^\dagger(b+k)$. Then

$$\frac{\|h\|_2}{\|x\|_2} \le \kappa\,\eta\,\frac{\|P_A k\|_2}{\|P_A b\|_2} \quad . \tag{5.2}$$

Proof. With the obvious partitioning of k we have $h = A_{11}^{-1}k_1$, so that

$$\|h\|_2 \le \|A_{11}^{-1}\|\|k_1\| \ . \tag{5.3}$$

But $\|x\|_2 = \eta^{-1}\|b_1\|_2/\|A\|_2$, which combined with (5.3) yields (5.2). □

Theorem 5.1 shows that the perturbation in x is determined by the projection of k onto $R(A)$. However, $P_A k$ is normalized by $\|P_A b\|_2$, and if this latter quantity is small, the perturbation may be large. Since

$$\|b\|_2^2 = \|P_A b\|_2^2 + \|r\|_2^2 \ ,$$

this observation may be summarized by saying that large residuals are troublesome, a statement which will be amply supported later.

Since η can be as small as κ^{-1}, the number κ cannot be taken as a condition number for perturbations in b without further qualification. If x does not reflect the ill-conditioning of A, then η is near unity and κ is a condition number. Otherwise the solution will be relatively insensitive to perturbations in b.

We next turn to assessing the effects on x of a perturbation in A.

<u>Theorem</u> 5.2. Let $x = A^{\dagger}b$ and $x + h = B^{\dagger}b$, where $B = A + E$ is an acute perturbation of A. Then

$$\frac{\|h\|_2}{\|x\|_2} \le \frac{\kappa}{\gamma}\frac{\|E_{11}\|_2}{\|A\|_2} + \psi_2\left(\frac{\kappa}{\gamma}\frac{\|E_{12}\|_2}{\|A\|_2}\right) + \frac{\kappa^2}{\gamma^2}\frac{\|E_{21}\|_2}{\|A\|_2}\left(\eta\frac{\|b_2\|_2}{\|b_1\|_2} + \frac{\|E_{21}\|_2}{\|A\|_2}\right) . \tag{5.4}$$

<u>Proof</u>. Write

$$h = J_{12}^{\dagger}(B_{11}^{-1} - A_{11}^{-1})b_1 + (J_{12}^{\dagger} - I_{12}^{\dagger})A_{11}^{-1}b_1 + J_{12}^{\dagger}B_{11}^{-1}(J_{21}^{\dagger} - I_{21}^{\dagger})b . \tag{5.5}$$

Then

$$\|J_{12}^{\dagger}(B_{11}^{-1} - A_{11}^{-1})b_1\|_2 \le \frac{\kappa}{\gamma}\frac{\|E_{11}\|_2}{\|A\|_2}\|x\|_2 , \tag{5.6}$$

and

$$\|(J_{12}^{\dagger} - I_{12}^{\dagger})A_{11}^{-1}b_1\|_2 \le \psi_2\left(\frac{\kappa}{\gamma}\frac{\|E_{12}\|_2}{\|A\|_2}\right)\|x\|_2 . \tag{5.7}$$

Now

$$\begin{aligned} J_{12}^{\dagger}B_{11}^{-1}(J_{21}^{\dagger} - I_{21}^{\dagger})b &= J_{12}^{\dagger}B_{11}^{-1}[(I + F_{21}^{H}F_{21})^{-1} - I]b_1 \\ &\quad + J_{12}^{\dagger}B_{11}^{-1}(I + F_{21}^{H}F_{21})^{-1}F_{21}^{H}b_2 . \end{aligned} \tag{5.8}$$

To bound the first term in (5.8), note that $(I+F_{21}^{H}F_{21})^{-1} - I = -(I+F_{21}^{H}F_{21})^{-1}F_{21}^{H}F_{21}$. Hence

$$
(5.9)\qquad
\begin{aligned}
&\|J_{12}^{\dagger}B_{11}^{-1}[(I+F_{21}^{H}F_{21})-I]b_1\|_2\\
&\quad\le \|B_{11}^{-1}\|_2\|(I+F_{21}^{H}F_{21})^{-1}\|_2\|F_{21}^{H}\|_2\|F_{21}b_1\|_2\\
&\quad\le \|B_{11}^{-1}\|_2^2\|E_{21}\|_2\|E_{21}B_{11}^{-1}b_1\|_2\\
&\quad\le \|B_{11}^{-1}\|_2^2\|E_{21}\|_2^2\|x\|_2 \le \left[\frac{\kappa}{\gamma}\frac{\|E_{21}\|_2}{\|A\|_2}\right]^2\|x\|_2 .
\end{aligned}
$$

For the second term in (5.8) we have

$$
(5.10)\qquad
\begin{aligned}
&\|J_{21}^{\dagger}B_{11}^{-1}(I+F_{21}^{H}F_{21})^{-1}F_{21}b_2\|_2\\
&\quad\le \|B_{11}^{-1}\|^2\|E_{21}\|\|b_2\|\\
&\quad= \|B_{11}^{-1}\|_2^2\|E_{21}\|_2\frac{\|b_2\|_2}{\|b_1\|_2}\eta\frac{\|x\|_2}{\|A\|_2}\\
&\quad\le \eta\left(\frac{\kappa}{\gamma}\right)^2\frac{\|E_{21}\|_2}{\|A\|_2}\frac{\|b_2\|}{\|b_1\|}\|x\|_2 .
\end{aligned}
$$

The bound (5.4) follows on combining (5.5)-(5.10). □

The first two terms in (5.4) are unexceptionable. The first term corresponds to the classical result for linear systems and is the only nonzero term when A is square and nonsingular. The second term depends on $P_A E R_A^{\perp}$ and vanishes when A is of full column rank, as it is in many applications.

The third term requires more explanation. If terms of second order in $\|E_{21}\|$ are ignored, this expression becomes essentially

$$
(5.11)\qquad \frac{\kappa^2}{\gamma^2}\eta\frac{\|b_2\|_2}{\|b_1\|_2}\frac{\|E_{21}\|_2}{\|A\|_2} \equiv \frac{\kappa^2}{\gamma^2}\eta\tan\theta\frac{\|E_{21}\|_2}{\|A\|_2} ,
$$

where θ is the angle subtended by b and $\mathcal{R}(A)$. The number $\kappa^2\eta \tan\theta/\gamma^2$ can vary from 0 to ∞. It is small when θ is small (i.e. the residual vector is small). It is also reduced in size when x reflects the ill-conditioning of A so that $\eta \cong \kappa^{-1}$. When x does not reflect the ill-conditioning of A and θ is significant, it is of order κ^2, thus making the third term in (5.4) the dominant one.

We have bounded the third term in the decomposition (5.5) in such a way as to reflect its behavior when E_{21} is small. In fact it is bounded for all values of E_{21}, and the third term in (5.4) may be replaced by

$$\frac{\kappa}{\gamma}\,\eta\,\frac{\|b\|_2}{\|b_1\|_2}\,\psi_2\left(\frac{\kappa}{\gamma}\,\frac{\|E_{21}\|_2}{\|A\|_2}\right).$$

The residual. Since the residual vector is given by $r = P_A b$, the theory of §4 may be applied to give perturbation bounds for the residual. Specifically, if

$$\hat{x} = B^{\dagger} b$$

and

$$\hat{r} = b - B\hat{x} = P_B b\ ,$$

then

$$\|\hat{r} - r\|_2 \le \|P_B - P_A\|_2 \|b\|_2$$

and $\|P_B - P_A\|_2$ can be bounded by (4.1) in Theorem 4.1.

In applications one may not be interested in $\hat{r}$; rather one is interested in the residual $\bar{r}$ of $\hat{x}$ with respect to the matrix A:

$$\bar{r} = b - A\hat{x} .$$

If we write

$$\bar{r} - r = (P_B - P_A)b - E\hat{x} ,$$

then

$$\|\bar{r}-r\|_2 \le \|P_B-P_A\|_2\|b\|_2 + \|E\|_2\|\hat{x}\|_2 .$$

Theorem 5.1 provides the necessary estimate of $\hat{x}$.

If we concern ourselves with only the change in $\|r\|_2$ we can derive a slightly stronger result. Since r is the minimizing residual, we have $\|r\|_2 \le \|\bar{r}\|_2$. Likewise $\|b-(A+E)\hat{x}\|_2 \le \|b-(A+E)x\|_2$, from which it follows that

$$\|r\|_2 \le \|\bar{r}\|_2 \le \|r\|_2 + \|E\|_2(\|x\|_2+\|\hat{x}\|_2) .$$

<u>Asymptotic forms and derivatives</u>. An asymptotic form for the perturbed least squares solution $\hat{x}$ can be obtained from (3.4):

$$\begin{aligned} \hat{x} = x - A^{\dagger}P_A E R_A x - R_A^{\perp}E^H P_A (A^H)^{\dagger} x \\ + (A^H A)^{\dagger} R_A E^H P_A^{\perp} b + O(\|E\|^2) . \end{aligned} \tag{5.12}$$

An equivalent asymptotic formula, which may be useful in computational work, can be derived from the reduced form (3.23). The derivative formula corresponsing to (5.12) is

$$\frac{dx}{d\tau} = -A^{\dagger}P_A \frac{dA}{d\tau} R_A x - R_A^{\perp} \frac{dA^H}{d\tau} P_A (A^H)^{\dagger} x + (A^H A)^{\dagger} R_A \frac{dA^H}{d\tau} P_A b \ .$$

An inverse perturbation theorem. Theorem 5.1 shows how a perturbation in A can affect the least squares solution. Here we consider the question: given a vector $\hat{x}$ under what conditions is $\hat{x}$ the least squares solution of a slightly perturbed problem? One such condition is given in the following theorem.

Theorem 5.2. Let $\hat{x} \in \mathbb{C}^n$ be given. Let $x = A^{\dagger}b$, $r = b - Ax$, and $\hat{r} = b - A\hat{x}$. If

$$\|\hat{r}\|_2^2 = \|r\|_2^2 + \varepsilon^2 ,$$

then there is a matrix E of rank unity with

$$\|E\|_2 = \frac{\varepsilon}{\|\hat{x}\|_2} \tag{5.13}$$

such that $\|b-(A+E)\hat{x}\|_2$ is a minimum.

Proof. Let

$$e = \hat{r} - r = (x-\hat{x}) \in R(A) \ .$$

Since $r \in R(A)^{\perp}$,

$$\|\hat{r}\|_2^2 = \|r\|_2^2 + \|e\|_2^2 ,$$

which shows that $\|e\|_2 = \varepsilon$. Let

$$E = -\frac{e\hat{x}^H}{\|\hat{x}\|_2^2}\,.$$

Then E satisfies (5.13) and $R(E) \subset R(A)$. Hence $R(A+E) \subset R(A)$. But

$$b - (A+E)\hat{x} = r \in R(A)^{\perp},$$

which shows that the residual $b - (A+E)\hat{x} \in R(A+E)^{\perp}$, and $\hat{x}$ solves the required least squares problem. □

A consequence of this theorem is that there is little use hunting for the exact minimizing x. Provided the residual is nearly minimal, the approximate solution $\hat{x}$, however inaccurate, is the exact solution of a slightly perturbed problem.

It is sometimes desirable that the perturbation matrix E in Theorem 5.2 not alter some of the columns of A (e.g. a column may be dates in years). This can be done as follows. Let $\tilde{x}$ be the vector obtained from $\hat{x}$ by setting to zero the components corresponding to the columns that are not to be disturbed. Then

$$\tilde{E} = -\frac{e\tilde{x}^H}{\|\tilde{x}\|_2^2}$$

is the required matrix. Of course $\|\tilde{x}\|_2 \le \|x\|_2$ so that $\|\tilde{E}\|_2 \ge \|E\|_2$; however $\|\tilde{E}\|_2$ may still be small enough for practical purposes.

<u>Notes and references</u>. Much of the perturbation theory for pseudo-inverses has been a byproduct of the search for bounds for the linear least squares problem. Golub and Wilkinson (1966) gave a first order analysis

of the problem and were the first to note the dependence of the solution on κ^2. Rigorous upper bounds were derived by Hanson and Lawson (1969), Pereyra (1969), and Stewart (1969). Wedin (1969) also gives bounds. More recent treatments have been given by Lawson and Hanson (1974) and Abdelmalek (1974). Van der Sluis (1975) was the first to point out the mitigating effect of η in (5.11).

The inverse perturbation theorem is new.

References

N. N. Abdelmalek (1974), On the solution of the linear least squares problem and pseudo-inverses, Computing 13, 215-228.

S. N. Afriat (1957), Orthogonal and oblique projectors and the characteristics of pairs of vector spaces, Proc. Cambridge Philos. Soc. 53, 800-816.

A. Ben-Israel (1966), On error bounds for generalized inverses, SIAM J. Numer. Anal. 3, 585-592.

__________ and T. N. E. Greville (1974), Generalized Inverses: Theory and Applications, John Wiley, New York.

T. L. Boullion and P. L. Odell (1971), Generalized Inverse Matrices, John Wiley, New York.

Chandler Davis and W. M. Kahan (1970), The rotation of eigenvectors by a perturbation - III, SIAM J. Numer. Anal. 7, 1-46.

I. C. Gohberg and M. G. Krein (1969), Introduction to the Theory of Nonselfadjoint Operators, Amer. Math. Soc., Providence, Rhode Island.

G. H. Golub (1965), Numerical methods for solving linear least squares problems, Numer. Math. 7, 206-216.

__________ and J. H. Wilkinson (1966), Note on the iterative refinement of least squares solution, Numer. Math. 9, 139-148.

__________ and V. Pereyra (1973), The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate, SIAM J. Numer. Anal. 10, 413-432.

__________ (1975), Differentiation of pseudoinverses, separable nonlinear least squares problems, and other tales, manuscript.

R. J. Hanson and C. L. Lawson (1969), Extensions and applications of the Householder algorithm for solving linear least squares problems, Math. Comp. 23, 787-812.

J. Z. Hearon and J. W. Evans (1968), Differentiable generalized inverses, J. Res. Nat. Bur. Stand. 72B, 109-113.

A. S. Householder (1964), The Theory of Matrices in Numerical Analysis, Dover, New York.

T. Kato (1966), Perturbation Theory for Linear Operators, Springer, Berlin.

C. L. Lawson and R. J. Hanson (1974), Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, New Jersey.

L. Mirsky (1960), Symmetric gauge functions and unitarily invariant norms, Quart. J. Math. Oxford (2) 11, 50-59.

J. von Neumann (1937), Some matrix-inequalities and metrization of matric-space, Tomsk. Univ. Rev. 1, 286-300.

M. Pavel-Parvu and A. Korganoff (1969), Iteration functions for solving polynomial equations, in Constructive Aspects of the Fundamental Theorem of Algebra (B. Dejon and P. Henrici, eds.), John Wiley, New York.

R. Penrose (1955), A generalized inverse for matrices, Proc. Cambridge Philos. Soc. 51, 506-513.

________ (1956), On best approximate solution of linear matrix equations, Proc. Cambridge Philos. Soc. 52, 17-19.

V. Pereyra (1969), Stability of general systems of linear equations, Aequat. Math. 2, 194-206.

C. R. Rao and S. K. Mitra (1971), Generalized Inverse of Matrices and its Applications, John Wiley, New York.

A. van der Sluis (1975), Stability of the solutions of linear least squares problems, Numer. Math. 23, 241-254.

G. W. Stewart (1969), On the continuity of the generalized inverse, SIAM J. Appl. Math 17, 33-45.

________ (1973), Introduction to Matrix Computations, Academic Press, New York.

P.-Å. Wedin (1969), On pseudo-inverses of perturbed matrices, Lund. Univ. Comput. Sci. TR.

________ (1973), Perturbation theory for pseudo-inverses, BIT 13, 217-232.

J. H. Wilkinson (1965), The Algebraic Eigenvalue Problem, Oxford, London.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER N00014-67-A-0218-0018-387	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 9
4. TITLE (and Subtitle) 6 ON THE PERTURBATION OF PSEUDO-INVERSES, PROJECTIONS AND LINEAR LEAST SQUARES PROBLEMS.		5. TYPE OF REPORT & PERIOD COVERED Technical Report,
		6. PERFORMING ORG. REPORT NUMBER Technical Report TR-387
7. AUTHOR(s) 10 G. W. Stewart 14 TR-387		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-67-A-0218-0018
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Science Department University of Maryland College Park, Maryland 20742		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Mathematics Branch Arlington, Virginia 22217		12. REPORT DATE 11 July 1975
		13. NUMBER OF PAGES 49
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12 50p.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

pseudoinverse
generalized inverse
least squares
projection
perturbation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper surveys perturbation theory for the pseudo-inverse (Moore-Penrose generalized inverse), for the orthogonal projection onto the column space of a matrix, and for the linear least squares problem.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 409022

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LB